Two Theorems Generalizing the Mean Transition Probability Results in the Theory of Markov Chains

D. J. Hartfiel
Mathematics Department
Texas A&M University
College Station, Texas 77843

Recommended by Hans Schneider

ABSTRACT

This paper gives two theorems which generalize the classical results of Markov chain theory concerning mean transition probabilities.

In considering the theory of Markov chains, two matrix results come to mind.

(1) If A is fully regular and stochastic, then $\lim_{k\to\infty}A^k=A_0$ exists and is rank one [1, p. 89].

(2) If A is stochastic with 1 a simple root of the characteristic equation of A, then

$$\lim_{k\to\infty}\frac{A+A^2+\cdots+A^k}{k}=A_0$$

exists and is rank one [1, p. 96].

The usual interpretation of these results concerns final transition probabilities of a Markov chain in which the matrix of transition probabilities at each step is given by A. In [2], [4], and [5] result (1) is generalized by considering $\lim_{k\to\infty}A_1A_2\cdots A_k$, where each A_i is stochastic. Conditions are laid down to guarantee that this limit exists and is rank one. This paper, then, is concerned with result (2)—in particular, with what condition can be imposed to guarantee that

$$\lim_{k\to\infty}\frac{A_1+A_1A_2+\cdots+A_1A_2\cdots A_k}{k}$$

exists and is rank one. Some related work of this nature may be found in [6].

For the reading of this paper, some familiarity with the classical theory of Markov chains, as given in [1], is necessary. Also, for nonnegative matrices of order n, we utilize

$$u(A)=\min_{|R|+|C|=n}\left(\max_{\substack{i\in R\\ j\in C}} a_{ij}\right),$$

where R and C denote nonempty subsets of row and column indices, respectively, with $|S|$ being the number of elements in set S.

RESULTS

The major tools in the arguments of the two theorems are as follows.

LEMMA 1. *Suppose A is stochastic and of order n. Let $x=(x_1,x_2,\dots,x_n)^t \geqslant 0$ and $Ax=z=(z_1,z_2,\dots,z_n)^t$. Then*

$$\min_i x_i \leqslant z_j \leqslant \max_i x_i \qquad \text{for } j=1,2,\dots,n.$$

LEMMA 2. *If B_1, $B_2,\dots,B_h$ are nonnegative matrices of order n and $u(B_k)\geqslant u$ for $k=1,2,\dots,h$, then $u(B_1B_2\cdots B_h)\geqslant u^h$* [3, *essentially the proof of Lemma* 3].

LEMMA 3. *Suppose A_1, $A_2,\dots,A_k,\dots$ is a sequence of stochastic matrices of order n such that*

$$\lim_{k\to\infty} A_k=A_0 \tag{1}$$

and

$$u(A_0)\geqslant u>0. \tag{2}$$

Then $\lim_{k\to\infty}A_1A_2\cdots A_k=A_0^\infty$, a rank one stochastic matrix [2].

Our first version of a generalized mean transition probability result is as follows.

THEOREM 1. Suppose A_1, $A_2,\dots,A_k,\dots$ is a sequence of stochastic matrices of order n. Let h be a positive integer such that

$$\lim_{k\to\infty} A_{kh+q}=B_q \tag{1}$$

and

$$u(B_q) \geqslant u > 0 \qquad \text{for } q = 0, 1, 2, \ldots, h-1. \tag{2}$$

Then

$$\lim_{k\to\infty} \frac{A_1 + A_1A_2 + \cdots + A_1A_2\cdots A_k}{k}$$
$$= \frac{(B_1B_2\cdots B_0)^{\infty} + (B_2B_3\cdots B_0B_1)^{\infty} + \ldots + (B_0B_1\cdots B_{h-1})^{\infty}}{}$$

Proof. Let $\varepsilon > 0$. From the hypothesis we have that

$$\lim_{k\to\infty} A_{kh+q+1}A_{kh+q+2}\cdots A_{kh+q+h} = B_{q+1}B_{q+2}\cdots B_{q+h}$$
$$\text{for } q = 0, 1, \ldots, h-1,$$

where the subscripts on the B's are $\operatorname{mod} h$. Further, by Lemma 2, $u(B_{q+1}B_{q+2}\cdots B_{q+h}) \geqslant u^h > 0$. Thus, from Lemma 3, we have that

$$\lim_{k\to\infty} (A_{h+q+1}A_{h+q+2}\cdots A_{h+q+h})$$
$$\times (A_{2h+q+1}A_{2h+q+2}\cdots A_{2h+q+h})\cdots(A_{kh+q+1}A_{kh+q+2}\cdots A_{kh+q+h})$$
$$= (B_{q+1}B_{q+2}\cdots B_{q+h})^{\infty}.$$

Pick N_1 so that for $k \geqslant N_1$,

$$\|(A_{h+q+1}A_{h+q+2}\cdots A_{h+q+h})\cdots(A_{kh+q+1}A_{kh+q+2}\cdots A_{kh+q+h})$$
$$-(B_{q+1}B_{q+2}\cdots B_{q+h})^{\infty}\| < \frac{\varepsilon}{3h},$$

where $\|A\| = \max_{i,j}|a_{ij}|$. Now as $(B_{q+1}B_{q+2}\cdots B_{q+h})^{\infty}$ is rank one and stochastic, it follows from Lemma 1 that

$$\|A_1A_2\cdots A_{(k+1)h+q} - (B_{q+1}B_{q+2}\cdots B_{q+h})^{\infty}\| \leqslant \frac{\varepsilon}{3h}.$$

Pick N_2 so that for $k \geqslant N_2$,

$$\frac{(N_1+1)h}{(k+1)h+q} < \frac{\varepsilon}{6} \qquad \text{for } q = 0, 1, \ldots, h-1.$$

Then for $k > \max\{N_1, N_2\}$ it follows that

$$\left\| \frac{A_1 + A_1A_2 + \cdots + A_1A_2 \cdots A_{(N_1+1)h}}{(k+1)h+q} \right.$$

$$+ \frac{A_1A_2 \cdots A_{(N_1+1)h+1} + \cdots + A_1A_2 \cdots A_{(k+1)h+q}}{(k+1)+q}$$

$$\left. - \frac{(B_1B_2 \cdots B_h)^\infty + (B_2B_3 \cdots B_hB_1)^\infty + \cdots + (B_hB_1 \cdots B_{h-1})^\infty}{h} \right\|$$

$$< \frac{\varepsilon}{6} + \left\| \frac{A_1A_2 \cdots A_{(N_1+1)h+1} + \cdots + A_1A_2 \cdots A_{(k+1)h+q}}{(k+1)h+q} \right.$$

$$\left. - \frac{(B_1B_2 \cdots B_h)^\infty + \cdots + (B_hB_1 \cdots B_{h-1})^\infty}{h} \right\|$$

and by regrouping,

$$< \frac{\varepsilon}{6} + \sum_{t=1}^{q} \left\| \frac{A_1A_2 \cdots A_{(N_1+1)h+t} + \cdots + A_1A_2 \cdots A_{(k+1)h+t}}{(k+1)h+q} \right.$$

$$\left. - \frac{(B_{t+1}B_{t+2} \cdots B_t)^\infty}{h} \right\|$$

$$+ \sum_{t=q+1}^{h} \left\| \frac{A_1A_2 \cdots A_{(N_1+1)h+t} + \cdots + A_1A_2 \cdots A_{kh+t}}{(k+1)h+q} - \frac{(B_{t+1}B_{t+2} \cdots B_t)^\infty}{h} \right\|$$

$$< \frac{\varepsilon}{6} + \sum_{t=1}^{q} \left\| \frac{(k+1)-N_1}{(k+1)h+q} (B_{t+1}B_{t+2} \cdots B_t)^\infty - \frac{(B_{t+1}B_{t+2} \cdots B_t)^\infty}{h} \right\|$$

$$+ \sum_{t=q+1}^{h} \left\| \frac{k-N_1}{(k+1)h+q} (B_{t+1}B_{t+2} \cdots B_t)^\infty - \frac{(B_{t+1}B_{t+2} \cdots B_t)^\infty}{h} \right\| + \frac{\varepsilon}{3}$$

$$< \frac{\varepsilon}{2} + \sum_{t=1}^{q} \left| \frac{k+1-N_1}{(k+1)h+q} - \frac{1}{h} \right| + \sum_{t=q+1}^{h} \left| \frac{k-N_1}{(k+1)h+q} - \frac{1}{h} \right|$$

$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \qquad \text{for } k > N_3, \quad \text{for some } N_3 > \max\{N_1, N_2\}.$$

As ε was arbitrary, the result follows. ∎

Our second version of a generalized mean transition probability result is as follows.

THEOREM 2. *Let P be a stochastic matrix with 1 a simple root of the characteristic equation of P, and P^h regular for some positive integer h. Suppose $A_1, A_2, \ldots, A_k, \ldots$ is a sequence of stochastic matrices of order n such that $\lim_{k\to\infty} A_k = P$. Then*

$$\lim_{k\to\infty} \frac{A_1 + A_1A_2 + \cdots + A_1A_2\cdots A_k}{k} = \frac{(I+P+\cdots+P^{h-1})P^{h\infty}}{h}.$$

Proof. Set $\lim_{k\to\infty} P^{hk+q} = Y_q$ for $q=0, 1,\ldots,h-1$. Let $\varepsilon>0$. Pick N_1 so that for $k \geqslant N_1$, $\|P^{hk+q} - Y_q\| \leqslant \varepsilon/4$ for $q=0, 1,\ldots,h-1$. Pick N_2 so that for $k \geqslant N_2$, $\|A_{k+1}\cdots A_{k+hN_1+q} - P^{hN_1+q}\| \leqslant \varepsilon/4$ for $q=0, 1,\ldots,h-1$. Thus $\|A_{k+1}\cdots A_{k+hN_1+q} - Y_q\| \leqslant \varepsilon/2$ for $q=0,\cdot,\ldots,h-1$. Therefore

$$\begin{aligned} &Y_0 + Y_1 + \cdots + Y_{h-1} - \frac{h\varepsilon}{2}J \\ &\leqslant A_kA_{k+1}\cdots A_{k+hN_1} + \cdots \\ &+ A_kA_{k+1}\cdots A_{k+hN_1+h-1} \\ &\leqslant Y_0 + Y_1 + \cdots + Y_{h-1} + \frac{h\varepsilon}{2}J, \end{aligned}$$

where J is the matrix of order n which has all of its entries equal to one. As $(Y_0 + Y_1 + \cdots + Y_{h-1})/h$ is rank one and stochastic,

$$\begin{aligned} Y_0 + Y_1 + \cdots + Y_{h-1} - \frac{h\varepsilon}{2}J \leqslant A_1A_2\cdots A_kA_{k+1}\cdots A_{k+hN_1} + \cdots \\ + A_1A_2\cdots A_kA_{k+1}\cdots \\ A_{(k+hN_1)+h-1} \leqslant Y_0 + Y_1 + \cdots + Y_{h-1} + \frac{h\varepsilon}{2}J. \end{aligned}$$

Thus

$$\begin{aligned} t(Y_0 + Y_1 + \cdots + Y_{h-1}) - \frac{th\varepsilon}{2}J - qJ \leqslant A_1A_2\cdots A_{N_2+hN_1} + \cdots \\ + A_1A_2\cdots A_{N_2+hN_1+th+q} \leqslant t(Y_0 + Y_1 + \cdots + Y_{h-1}) + \frac{th\varepsilon}{2}J + qJ \end{aligned}$$

$$\text{for } t \geqslant 1 \quad \text{and} \quad q=0, 1,\ldots,h-1.$$

Therefore,

$$t(Y_0+Y_1+\cdots+Y_{h-1})-\frac{th\varepsilon}{2}J-qJ-(N_2+hN_1)J$$

$$\leqslant A_1+A_1A_2+\cdots+A_1A_2\cdots A_{N_2+hN_1}+\cdots+A_1A_2\cdots A_{N_2+hN_1+th+q}$$

$$\leqslant t(Y_0+Y_1+\cdots+Y_{h-1})+\frac{th\varepsilon}{2}J+qJ+(N_2+hN_1)J.$$

Set $R=N_2+hN_1+th+q$, and divide the above inequality by R. Then for R sufficiently large (i.e., t sufficiently large),

$$\frac{Y_0+Y_1+\cdots+Y_{h-1}}{h}-\varepsilon J\leqslant\frac{A_1+A_1A_2+\cdots+A_1A_2\cdots A_R}{R}$$

$$\leqslant\frac{Y_0+Y_1+\cdots+Y_{h-1}}{h}+\varepsilon J.$$

As ε was arbitrary,

$$\lim_{R\to\infty}\frac{A_1+A_1A_2+\cdots+A_1A_2\cdots A_R}{R}=\frac{Y_0+Y_1+\cdots+Y_{h-1}}{h},$$

and the result follows. ■

As an indication of some necessity for the hypothesis of Theorem 1 and Theorem 2, we provide the following example.

EXAMPLE. Let

$$P=\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0\\ 0 & 0 & 1 & \cdots & 0 & 0\\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot\\ 0 & 0 & 0 & \cdots & 0 & 1\\ 1 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

be of order n. Define a sequence $P_1, P_2,\ldots$ of stochastic matrices of order n so that

$$P_k=\begin{cases} P & \text{or}\\ I & \text{or},\\ P^t \end{cases}$$

with

$$P_1 = P_1P_2 = I,$$
$$P_1P_2P_3 = \cdots = P_1P_2P_3P_4 = P,$$
$$\cdots$$
$$P_1P_2\cdots P_{2^k+1} = \cdots = P_1P_2\cdots P_{2^{k+1}} = I,$$
$$P_1P_2\cdots P_{2^{k+1}+1} = \cdots = P_1P_2\cdots P_{2^{k+2}} = P,$$

for k even. Then $(P_1 + P_1P_2 + \cdots + P_1P_2\cdots P_R)/R$ does not converge.

REFERENCES

1 F. R. Gantmacher, *The Theory of Matrices*, Vol. 2, Chelsea, New York, 1962.
2 D. J. Hartfiel, A result concerning strongly erogodic nonhomogeneous Markov chains, *Linear Algebra Appl.* **9** (1975), 169–174.
3 D. J. Hartfiel, On infinite products of nonnegative matrices, *SIAM J. Appl. Math.* **26** (1974), 824–828.
4 J. Hajnal, Weak ergodicity in non-homogeneous Markov chains, *Proc. Camb. Philos. Soc.* **54** (1958), 233–246.
5 J. L. Mott, Conditions for the ergodicity of non-homogeneous finite Markov chains, *Proc. Roy. Soc. Edinb.* **64** (1957), 369–380.
6 J. L. Mott, The central limit theorem for a convergent nonhomogeneous finite Markov chain, *Proc. Roy. Soc. Edinb.* **65** (1958), 109–120.
7 A. Paz, *Introduction to Probabilistic Automata*, Academic, New York, 1971.

Received July 25, 1974